

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号
特表2001-511546
(P2001-511546A)

(43) 公表日 平成13年8月14日 (2001.8.14)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 0 6 F 17/30	1 7 0	G 0 6 F 17/30	1 7 0 F
C 1 2 Q 1/68		C 1 2 Q 1/68	A
// C 1 2 M 1/00		C 1 2 M 1/00	A
G 0 6 F 159:00		G 0 6 F 159:00	

審査請求 未請求 予備審査請求 有 (全 39 頁)

(21) 出願番号	特願2000-504290(P2000-504290)	(71) 出願人	アフメトリックス インコーポレイテッド
(86) (22) 出願日	平成10年7月24日(1998.7.24)		アメリカ合衆国 カリフォルニア州
(85) 翻訳文提出日	平成12年1月25日(2000.1.25)		95051 サンタ クララ セントラル エクスプレスウェイ 3380
(86) 国際出願番号	PCT/US 98/15151	(72) 発明者	バラバン, デイビッド ジェイ.
(87) 国際公開番号	WO 99/05323		アメリカ合衆国 カリフォルニア 94901,
(87) 国際公開日	平成11年2月4日(1999.2.4)		サン ラファエル, プレット ハート
(31) 優先権主張番号	60/053, 842		ロード 37
(32) 優先日	平成9年7月25日(1997.7.25)	(74) 代理人	弁理士 大塩 竹志
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/069, 198		
(32) 優先日	平成9年12月11日(1997.12.11)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 遺伝子発現および評価システム

(57) 【要約】

遺伝子発現データベースのための照会システムを効率的かつ容易に使用すること。このようなシステムを使用することによって、その発現が特定の組織タイプに相関する遺伝子または発現配列タグを容易に同定し得る。種々の組織タイプは、異なる疾患、疾患の進行の状態、異なる器官、異なる種などに対応し得る。研究者らは、今や十分有利に大規模な遺伝子発現データベースを使用し得る。

【特許請求の範囲】

【請求項1】 コンピュータシステムにおいて、発現レベル情報を蓄積するデータベースを操作する方法であって、以下：

複数の組織タイプのそれぞれにおいて測定されるような、複数の遺伝子または発現配列タグ（EST）のそれぞれについての発現レベルを含むデータベースを提供する工程；

該データベースに対するユーザ照会を受納し、該複数の遺伝子またはESTの所望のものを同定する工程であって、該ユーザ照会は、該所望の遺伝子の発現レベル特性を特定する、工程；および

該発現レベル特性を、該データベースに蓄積された該発現レベルに対して比較し、該所望の遺伝子またはESTを同定する工程、を包含する、方法。

【請求項2】 前記所望の遺伝子またはESTを同定する情報を表示する工程をさらに含む、請求項1に記載の方法。

【請求項3】 前記複数の組織タイプが罹患組織タイプを含む、請求項1に記載の方法。

【請求項4】 前記複数の組織タイプが健常組織タイプを含む、請求項1に記載の方法。

【請求項5】 前記複数の組織タイプがガン性組織タイプを含む、請求項1に記載の方法。

【請求項6】 前記複数の組織タイプが、薬物処理された組織タイプを含む、請求項1に記載の方法。

【請求項7】 前記複数の組織タイプが、異なる種から得られた組織を含む、請求項1に記載の方法。

【請求項8】 前記複数の組織タイプが、異なる器官から得られた組織を含む、請求項1に記載の方法。

【請求項9】 前記発現レベル特性が、前記複数の組織タイプの少なくとも2つにおいて特定の遺伝子について測定されるような発現レベル範囲を含む、請求項1に記載の方法。

【請求項10】 前記発現レベル特性が、前記複数の組織タイプの少なくとも2つにおいて特定の遺伝子について測定されるような発現レベル間の関連性を含む、請求項1に記載の方法。

【請求項11】 請求項1に記載の方法であって、以下：

画面表示のために2つの前記複数の組織タイプを選択するユーザ入力を受納する工程；

該2つの組織タイプの第1の組織タイプに対応する第1の軸を表示する工程；

該2つの組織タイプの第2の組織タイプに対応する第2の軸を表示する工程；

前記複数の遺伝子またはESTの選択されたものについて、ある位置にある印を表示する工程であって、ここで該位置は、該第1の組織タイプにおいて測定される前記選択された遺伝子またはESTの発現レベルに従って該第1の軸に対して選択され、そして該第2の組織タイプにおいて測定される前記選択された遺伝子またはESTの発現レベルに従って該第2の軸に対して選択される、工程、をさらに包含する、方法。

【請求項12】 複数の選択された遺伝子またはESTについて、印を表示する前記操作を反復する工程をさらに含む、請求項11に記載の方法。

【請求項13】 コンピュータシステムにおいて、化合物濃度についての情報を蓄積するデータベースを操作する方法であって、以下：

複数のサンプルにおいて測定されるような、複数の化合物の濃度を含むデータベースを提供する工程；

該データベースに対するユーザ照会を受納し、該複数の化合物の所望のものを同定する工程であって、該ユーザ照会は、該複数のサンプルの選択されたサンプルにおける該所望の化合物の濃度特性を特定する、工程；および、

該濃度特性を、該データベースに蓄積された該濃度に対して比較し、該所望の化合物を同定する工程、を包含する、方法。

【請求項14】 発現レベル情報を蓄積するデータベースを操作するためのコンピュータプログラム製品であって、以下：

複数の組織タイプのそれぞれにおいて測定されるような、複数の遺伝子または

発現配列タグ (E S T) のそれぞれについての発現レベルを含むデータベースを提供するコード；

該データベースに対するユーザ照会を受納し、該複数の遺伝子またはE S Tの所望のものを同定するコードであって、該ユーザ照会は、該所望の遺伝子の発現レベル特性を特定する、コード；および

該発現レベル特性を、該データベースに蓄積された該発現レベルに対して比較し、該所望の遺伝子またはE S Tを同定するコード；および

コードを蓄積するためのコンピュータ読み出し可能な記憶媒体、を含む、製品。

【請求項15】 前記所望の遺伝子またはE S Tを同定する情報を表示するコードをさらに含む、請求項14に記載の製品。

【請求項16】 前記複数の組織タイプが罹患組織タイプを含む、請求項14に記載の製品。

【請求項17】 前記複数の健常タイプが健常組織タイプを含む、請求項14に記載の製品。

【請求項18】 前記複数の組織タイプがガン性組織タイプを含む、請求項14に記載の製品。

【請求項19】 前記複数の組織タイプが薬物処理した組織タイプを含む、請求項14に記載の製品。

【請求項20】 前記複数の組織タイプが、異なる種から得られた組織を含む、請求項14に記載の製品。

【請求項21】 前記複数の組織タイプが、異なる器官から得られた組織を含む、請求項14に記載の製品。

【請求項22】 前記発現レベル特性が、前記複数の組織タイプの少なくとも2つにおいて特定の遺伝子について測定されるような発現レベル範囲を含む、請求項14に記載の製品。

【請求項23】 前記発現レベル特性が、前記複数の組織タイプの少なくとも2つにおいて特定の遺伝子について測定されるような発現レベル間の関連性を含む、請求項14に記載の製品。

【請求項24】 請求項14に記載の製品であって、以下：

画面表示のために2つの前記複数の組織タイプを選択するユーザ入力を受納するコード；

該2つの組織タイプの第1の組織に対応する第1の軸を表示するコード；

該2つの組織タイプの第2の組織に対応する第2の軸を表示するコード；

前記複数の遺伝子またはESTの選択されたものについて、ある位置にある印を表示するコードであって、ここで該位置は、該第1の組織タイプにおいて測定される前記選択された遺伝子またはESTの発現レベルに従って該第1の軸に対して選択され、そして該第2の組織タイプにおいて測定される前記選択された遺伝子またはESTの発現レベルに従って該第2の軸に対して選択される、コード

をさらに含む、製品。

【請求項25】 複数の選択された遺伝子またはESTについて、印を表示する前記コードを反復的に適用するコードをさらに含む、請求項24に記載の製品。

【請求項26】 化合物濃度についての情報を蓄積するデータベースを操作するコンピュータプログラム製品であって、以下：

複数のサンプルにおいて測定されるような、複数の化合物の濃度を含むデータベースを受け取るコード；

該データベースに対するユーザ照会を受納し、該複数の化合物の所望のものを同定するコードであって、該ユーザ照会は、該複数のサンプルの選択されたものにおける該所望の化合物の濃度特性を特定する、コード；および、

該濃度特性を、該データベースに蓄積された該濃度に対して比較し、該所望の化合物を同定するコード、

を含む、製品。

【請求項27】 コンピュータシステムであって、以下：

プロセッサ；および

該プロセッサを操作するためのメモリー蓄積コードを含むシステムであって、ここで該コードは：

複数の組織タイプのそれぞれにおいて測定されるような、複数の遺伝子または発現配列タグ (E S T) のそれぞれについての発現レベルを含むデータベースを提供するコード ;

該データベースに対するユーザ照会を受納し、該複数の遺伝子または E S T の所望のものを同定するコードであって、該ユーザ照会は、該所望の遺伝子の発現レベル特性を特定する、コード ; および

該発現レベル特性を、該データベースに蓄積された該発現レベルに対して比較し、該所望の遺伝子または E S T を同定するコード、を含む、

システム。

【発明の詳細な説明】

【0001】

(関連出願の相互参照)

本願は、1997年7月25日に出願された、「COMPREHENSIVE BIO-INFORMATICS DATABASE」と題する米国仮出願第60/053,842号、1997年12月11日に出願された、「COMPREHENSIVE DATABASE FOR BIOINFORMATICS」と題する米国仮出願第60/069,198号、および1997年12月11日に出願された、「GENE EXPRESSION AND EVALUATION SYSTEM」と題する米国仮出願第60/069,436号からの優先権を主張している。3件全ての仮出願の内容は、本明細書中において参考として援用される。

【0002】

本願の主題は、本願と同日に出願された、同一人に譲渡される以下3件の出願の主題に関する。それらは「METHOD AND APPARATUS FOR PROVIDING A BIOINFORMATICS DATABASE」(代理人事件番号018547-033810号)、「METHOD AND SYSTEM FOR PROVIDING A POLYMORPHISM DATABASE」(代理人事件番号018547-033820号)、「METHOD AND SYSTEM FOR PROVIDING A PROBE ARRAY CHIP DESIGN DATABASE」(代理人事件番号018547-033830号)である。これら3件の出願の内容は、本明細書中において参考として援用される。

【0003】

(発明の背景)

本発明は、コンピュータシステム、およびより特定には、発現レベルまたは濃度を解析するためのコンピュータシステムに関する。

【0004】

多数の組織サンプルにおける遺伝子発現または発現配列タグ(EST)発現に

関する情報を収集するためのデバイスおよびコンピュータシステムが開発されてきている。例えば、PCT出願WO92/10588（本明細書中において、あらゆる目的のために参考として援用される）は、核酸および他の物質を配列決定するため、または配列検査するための技術を記載している。これらの作業を行うためのプローブは、例えば米国特許第5,143,854号および米国特許第5,571,639号（両件ともに本明細書中において、あらゆる目的のために参考として援用される）に開示されている開拓的技術の方法に従ってアレイにおいて形成され得る。

【 0 0 0 5 】

そこで記載されている技術の1つの局面によると、核酸プローブのアレイは、チップまたは基板上の既知の位置において製作される。次いで、蛍光標識化核酸は、チップと接触させられ、そしてスキャナーは標識化核酸がチップに結合する位置を示す画像ファイルを生成する。これらの位置におけるプローブの同一性に基づき、情報（例えば、DNAまたはRNAのモノマー配列）を抽出することが可能となる。

【 0 0 0 6 】

欧州特許公報第0848067号およびPCT公開公報第WO97/10365号（これらの内容は、本明細書中に参考として援用される）に開示されるような、このようなプローブアレイを使用して遺伝子発現をモニターするためのコンピュータ援用技術が、開発されている。多くの疾患状態は、遺伝子DNAのコピー数の変化または特定の遺伝子の転写レベル（例えば、開始の制御、RNA前駆体の提供、RNAプロセッシングなどによる）の変化のいずれかによる種々の遺伝子の発現レベルにおける差異によって特徴付けられる。例えば、遺伝子物質の欠損および獲得が、悪性形質転換および進行において重要な役割を果たしている。さらに、特定の遺伝子（例えば、癌遺伝子または腫瘍抑制遺伝子）の発現（転写）レベルにおける変化が、種々の癌の存在および進行の道標として働く。

【 0 0 0 7 】

遺伝子または発現配列タグの発現に関する情報は、多くの方法（上記のプローブアレイ技術を包含する）で、大規模に収集され得る。この情報の収集における

目的の1つは、その発現が特に重要である遺伝子またはESTの同定である。研究者らは、以下のような設問への回答を望む：1) どの遺伝子が悪性腫瘍の細胞において発現され、健常な組織または特定のレジメに従って処置された組織のいずれかにおいて発現されないか、2) どの遺伝子またはESTが、特定の器官で発現され、しかし他の器官では発現されないか、3) どの遺伝子またはESTが、特定の種で発現され、しかし他の種では発現されないか。

【0008】

上述の組織タイプの全てを含む多数のサンプルから莫大な量の発現データを収集することは、これらの設問への解答における第一歩でしかない。発現データの収集および蓄積においてなされた投資から十分な値を引き出すために、特に関連のある項目を見出すためにこれらデータを効率的に探知することができなければならない。必要とされることは、遺伝子発現データベースのための照会システムを効率的かつ容易に使用することである。

【0009】

(発明の要旨)

遺伝子発現データベースのための照会システムを効率的かつ容易に使用することが、本発明によって提供される。このようなシステムを使用して、当業者は、その発現が特定の組織タイプに相関する遺伝子または発現配列タグを、容易に同定し得る。種々の組織タイプは、異なる疾患、疾患進行の状態、異なる器官、異なる種などに対応し得る。研究者らは、現在では、大規模な遺伝子発現データベースを十分に有利に使用し得る。

【0010】

本発明の第一の局面に従って、コンピュータシステムにおいて、化合物濃度に関する情報を蓄積するデータベースを操作するための方法が、提供される。この方法は、以下の工程を包含する：複数のサンプルにおいて測定された複数の化合物の濃度を含むデータベースを提供する工程、このデータベースに対するユーザ照会を受納し、複数の化合物の所望のものを同定する工程であって、ここでこのユーザ照会は、複数のサンプルの選択されたものにおける所望の化合物の濃度特性を特定する、工程、およびこの濃度特性をデータベースに蓄積された濃度に対

して比較し、所望の化合物を同定する工程。

【 0 0 1 1 】

本明細書中の発明の性質および利点のさらなる理解は、本明細書の残りの部分および添付の図面を参照してなされ得る。

【 0 0 1 2 】

(特定の実施態様の説明)

図 1 は、本発明のソフトウェア実施態様を実行するために使用され得るコンピュータシステムの一例を図示する。図 1 は、モニタ 3、スクリーン 5、キャビネット 7、キーボード 9、およびマウス 11 を備えるコンピュータシステム 1 を示す。マウス 11 は、1 つ以上のボタン（例えば、マウスボタン 13）を有し得る。キャビネット 7 は、本発明を組み込むコンピュータコードを含むソフトウェアプログラムを格納および検索するために使用され得る CD-ROM ドライブ 15 およびハードドライブ（示さず）を収容する。CD-ROM 17 がコンピュータ読み出し可能媒体として示されるが、他のコンピュータ読み出し可能媒体（フロッピーディスク、DRAM、ハードドライブ、フラッシュメモリ、テープなどを含む）も利用され得る。キャビネット 7 はまた、通常のコンピュータコンポーネント（示さず）（例えば、プロセッサ、メモリなど）もまた収容する。

【 0 0 1 3 】

図 2 は、本発明のソフトウェア実施態様を実行するために使用されるコンピュータシステム 1 のシステムブロック図を示す。図 1 におけるように、コンピュータシステム 1 は、モニタ 3 およびキーボード 9 を備える。コンピュータシステム 1 はさらに、中央プロセッサ 50、システムメモリ 52、入力／出力（I/O）コントローラ 54、ディスプレイアダプタ 56、リムーバブルディスク 58、固定ディスク 60、ネットワークインターフェース 62、およびスピーカー 64 のようなサブシステムを備える。リムーバブルディスク 58 は、リムーバブルコンピュータ読み出し可能媒体（フロッピー、テープ、CD-ROM、リムーバブルハードドライブ、フラッシュメモリなどのような）を表す。固定ディスク 60 は、内蔵ハードドライブなどを表す。本発明との使用に適切な他のコンピュータシステムは、さらなるサブシステムまたはより少ないサブシステムを備え得る。例

えば、別のコンピュータシステムは、1つより多くのプロセッサ50（すなわち、マルチプロセッサシステム）またはメモリキャッシュを備え得る。

【0014】

矢印（例えば、66）は、コンピュータシステム1のシステムバスアーキテクチャを表す。しかし、これらの矢印は、サブシステムを連結するように働く任意の相互接続スキームを図示する。例えば、ディスプレイアダプタ56は、ローカルバスを介して中央プロセッサ50に接続され得るか、またはこのシステムは、メモリキャッシュを備え得る。図2に示したコンピュータシステム1は、本発明との使用に適切なコンピュータシステムの一例でしかない。本発明との使用に適切な他のサブシステム配置は、当業者に容易に明らかである。1つの実施態様では、このコンピュータシステムは、IBM互換性パーソナルコンピュータである。

【0015】

VLSIPSTMおよびGeneChipTM技術は、非常に小さなチップ上でポリマー（例えば、核酸）の非常に大きなアレイを作製および使用方法を提供する。米国特許第5,143,854号およびPCT特許公開第WO90/15070号および同第WO92/10092号（これらの各々は、あらゆる目的のためにこれにより参考として援用される）を参照のこと。チップ上の核酸プローブは、目的のサンプル核酸（「標的」核酸）において相補的核酸配列を検出するために使用される。

【0016】

プローブは核酸プローブである必要はなく、ペプチドのような他のポリマーでもあり得ることが理解されるべきである。ペプチドプローブは、サンプル中のペプチド、ポリペプチド、またはポリマーの濃度を検出するために使用され得る。測定に使用されるべきである濃度の化合物に対して結合親和性を有するプローブが、注意深く選択されるべきである。

【0017】

1つの実施態様では、本発明は、ポリマープローブのようなポリマーへの化合物の親和性をモニターすることによって測定されるような、サンプル中の化合物

の濃度に関する情報を検討および解析する方法を提供する。特定の適用では、濃度情報は、ハイブリダイズされた核酸プローブを含むチップについてのハイブリダイゼーション強度ファイルの解析によって生成される。特定のプローブへの核酸サンプルのハイブリダイゼーションは、さらに1つの遺伝子または発現配列タグ (E S T) の発現レベルを表し得る。本明細書中では、遺伝子または E S T の発現レベルは、サンプル内の、その遺伝子または E S T の転写から生じる m R N A またはタンパク質の濃度であると理解される。

【 0 0 1 8 】

本発明によって検討および／または解析される発現レベル情報は、プローブから得られる必要はなく、任意の起源から生じ得る。この発現情報がプローブアレイから収集される場合、プローブアレイは、サイズおよび密度について特定の基準を満たす必要はない。さらに、本発明は、ハイブリダイゼーションのような結合の蛍光測定を検討および／または解析すること限定されず、他の測定の検討および／または解析についても容易に利用され得る。

【 0 0 1 9 】

核酸以外の化合物の濃度は、本発明の1つの実施態様に従って検討および／または解析され得る。例えば、プローブアレイは、ペプチドプローブを含み得、これは、このペプチドプローブに結合するかもしれないかまたはしないかもしれないタンパク質サンプル、ポリペプチドサンプル、またはペプチドサンプルに曝露され得る。ペプチドプローブの適切な選択により、このペプチドプローブに結合する特定のタンパク質、ポリペプチド、またはペプチドの存在または非存在を検出し得る。

【 0 0 2 0 】

チップマスクを設計し、そのチップ上でプローブを合成し、標的サンプルからの核酸を標識し、そしてハイブリダイズしたプローブをスキャンするシステムが、米国特許第5, 571, 639号（これは、あらゆる目的のために、これにより参考として援用されている）に記載されている。しかし、本発明は、発現情報を生成するために他のシステムの結果を検討および／または解析するために、または核酸以外のポリマーの濃度を検討および／または解析するために、別個に使

用され得る。

【 0 0 2 1 】

用語「完全マッチプローブ」は、特定の標的配列に対して完全に相補的である配列を有するプローブをいう。試験プローブは、代表的には、標的配列の部分（サブ配列）に完全に相補的である。用語「ミスマッチコントロール」または「ミスマッチプローブ」は、その配列が、特定の標的配列に対して完全には相補的でないように故意に選択されるプローブを意味する。アレイ中の各ミスマッチ（M M）コントロールについて、代表的には、同じ特定の標的配列に対して完全に相補的である、対応する完全マッチ（P M）プローブが存在する。

【 0 0 2 2 】

チップから得られる情報の重要な断片のなかには、完全マッチプローブおよびミスマッチプローブから得られる相対的蛍光強度がある。これらの強度レベルは、遺伝子またはESTについての発現レベルを概算するために使用される。解析のために使用されるコンピュータシステムは、好ましくは、実験の利用可能な他の詳細（おそらく、遺伝子名、遺伝子配列、プローブ配列、基板上のプローブ位置などを含む）を有する。

【 0 0 2 3 】

発現解析は、各実験の各遺伝子について実施される。図3は、チップ上の特定の实验において測定されるような、特定の遺伝子について発現レベルを概算する工程を記載するフローチャートである。工程302で、このコンピュータシステムは、完全マッチプローブとミスマッチプローブとN対の生のスキャンデータを受納する。好ましい実施態様では、ハイブリダイゼーション強度は、基板上のプローブにハイブリダイスした蛍光標識化標的からの光子計数である。簡便化のために、完全マッチプローブのハイブリダイゼーション強度は「I₀」と称し、そしてミスマッチプローブのハイブリダイゼーション強度は「I₁」と称する。

【 0 0 2 4 】

工程304で、一対のプローブについてのハイブリダイゼーション強度を検索する。工程306で、バックグラウンドシグナル強度を各対の各ハイブリダイゼーション強度から差し引く。バックグラウンド減算はまた、全ての生のスキャン

データ上で同時に行われ得る。

【 0 0 2 5 】

工程 3 0 8 で、プローブの対のハイブリダイゼーション強度は、差閾値 (D) および比閾値 (R) と比較される。対のハイブリダイゼーション強度間の差 ($I_{..} - I_{..}$) が差閾値より大きいまたは等しいかどうか、および対のハイブリダイゼーション強度の商 ($I_{..} / I_{..}$) が比閾値より大きいまたは等しいかどうか決定される。差閾値は、代表的にはユーザにより決定される値であり、これは遺伝子 (単数または複数) の正確な発現モニタリングを生成するように決定されている。1つの実施態様において、差閾値は 2 0 であり、そして比閾値は 1 . 2 である。

【 0 0 2 6 】

$I_{..} - I_{..} \geq D$ かつ $I_{..} / I_{..} \geq R$ である場合、値 N P O S は、工程 3 1 0 で増加される。一般に、N P O S は、その遺伝子が発現されるらしいことを示すハイブリダイゼーション強度を有するプローブの対の数を示す値である。N P O S は、その遺伝子の発現の決定において利用される。

【 0 0 2 7 】

工程 3 1 2 で、 $I_{..} - I_{..} \geq D$ かどうか、および $I_{..} / I_{..} \geq R$ であるかどうかを決定する。これらの発現が真である場合、値 N N E G は、工程 3 1 4 で増加される。一般に、N N E G は、その遺伝子が発現されないらしいことを示すハイブリダイゼーション強度を有するプローブの対の数を示す値である。N N E G は、N P O S と同様に、遺伝子の発現の決定において利用される。

【 0 0 2 8 】

その遺伝子が発現されるか、または発現されないかのいずれかを示すハイブリダイゼーション強度を示す各々の対について、対数比值 (L R) および強度差値 (I D I F) は、工程 3 1 6 において計算される。L R は、その対のハイブリダイゼーション強度の商 ($I_{..} / I_{..}$) の対数によって計算される。I D I F は、対のハイブリダイゼーション強度間の差 ($I_{..} - I_{..}$) によって計算される。工程 3 1 8 で、ハイブリダイゼーション強度の次の対が存在する場合、それらは工程 3 0 4 において検索される。

【 0 0 2 9 】

行われる各分析について、特定のデータが、発現分析データベースに蓄積される。好ましくは、チップが発現を測定する各遺伝子またはESTについてのレコードが存在する。このレコードには、多様な情報を保持するようなフィールドが存在する。1つのフィールドは、分析を識別するための分析IDを蓄積する。結果タイプIDフィールドは、その列挙された発現結果が、その遺伝子が値P1、P2、P3、およびP4への決定マトリックスの適用に基づいて存在するか、ほとんどないか、存在しないか、または未知であることを示すか否かを示す。正のフィールドの数は、NPoSを示す。負のフィールドの数はNNEGを示す。使用されるフィールドの数は、NNEGまたはNPoSを増加させた対に帰属するプローブの数を示す。全てのフィールドの数は、Nを示す。平均の対数比フィールドは、全てのプローブ対についての平均のLRを示す。正の超過したフィールドの数は、NPoS - NNEGの値を示す。負の超過したフィールドの数は、NNEG - NPoSの値を示す。平均の差次的強度フィールドは、そのプローブ対についての平均のIDIFを示す。平均のフィールド中の数は、平均値を計算するのに使用されるプローブ対の数を示す。

【 0 0 3 0 】

発現データベースに対するユーザインターフェースを操作する工程は、ここで図4を参照して例示される。図4の工程は、繰り返され得るか、または異なる順序で生じ得るか、または1つ以上の工程が省略され得る。ユーザインターフェースの考察はまた、ユーザインターフェースの代表的なスクリーンディスプレイを示す図5A～5Lを参照する。

【 0 0 3 1 】

工程402で、ユーザは、照会のための発現分析結果のファイルを選択する。図5Aは、ユーザが発現結果ファイルを特定し得るインターフェーススクリーンを例示する。各ファイルは、1つの実験を表す。テーブル502は、すでに選択されたファイルを列挙する。得られたリストは、ボタン504を選択することによって後の使用のために保存され得る。以前に保存されたリストは、ボタン506を選択することによって削除され得る。ボタン508は、テーブル502に示

されるリストを以前に保存されたバージョンに対してリセットする。インポートボタン512は、テーブル502に示されるファイルの内容を、照会のためにインポートする。テーブル502内で、ファイル名列は、インポートボタン512の適用によってインポートされるファイル名を列挙する。コード列は、各ファイル中の発現データについての組織タイプを示す。複製ファイルはそのファイルが重複か否かを示す。チップ設計コード列は、そのファイルのデータを生成するために使用されたチップ設計を示す。種々の他の列（示さず）は、分析結果データについてのさらなる情報を与える。

【0032】

ファイル選択ボタン514を選択することによって、ユーザは、図5Bに示すように、選択されたファイルをスクリーン516に呼び出す。これは、インタラクティブなファイル検索、およびファイル名をタイプ入力することを必要としない選択プロセスを提供する。ファイルリストをインポートする前に、ユーザは、図5Cに示すように、種ドロップダウンリスト518を使用することによって種を選択するべきである。分析タイプドロップダウンリスト519は、相対的な発現分析と絶対的な発現分析との間をユーザが選択することを可能にする。

【0033】

図5Dは、工程404でインポートされた発現結果を正規化するための正規化フォーム520を示す。このソフトウェアは、正規化フォーム520上でのユーザの選択に基づく分析ルーチンによって生成される平均の差異データを測定する。チップ変動性の領域522において、ユーザは、ハウスキーピング遺伝子を公知の発現レベルで特定し、そしてスケール値を選択する。ユーザは、このスケール値を適用するかまたは適用しないかを選択することができる。ユーザがそのスケール値を適用することを選択した場合、単一のチップ上で測定された各遺伝子の発現レベルは、そのチップ上で測定されたハウスキーピング発現レベルの平均値で除算された所望のスケール因子に等しい値で積算される。

【0034】

また、正規化フォーム520上の組織変動性領域524において、ユーザは、多数のチップから回収されたデータに適用するスケール値、およびそれを適用す

るか否かを選択し得る。このスケール値が適用される場合、チップのセットにおいて測定された各発現値は、全チップセットについての全ての遺伝子にわたって測定された平均の発現レベルによって除算されたスケール値に等しい因子によって積算される。変換領域526は、負の平均の差異値が、対数変換の使用によって正の数に変換されるべきであるか否かをユーザが選択することを可能にする。ユーザは、リセットボタン528を選択することによって正規化フォーム520上でなされた全ての変化をリセットし得るか、または適用ボタン530を選択することによって選択された正規化および変換を適用し得る。

【 0 0 3 5 】

工程406で、ユーザは、インポート、正規化、および変換された実験データのより大きなセットをフィルタする。図5Eは、フィルタ実験フォーム532を示す。下側のテーブル534は、インポートされた実験、および遺伝子またはEST、ならびに実験および遺伝子またはESTの各組み合わせと関連する発現データを列挙する。上側のテーブル536は、下側のテーブル534における実験データをフィルタするための照会を入力するために使用される。上側のテーブル536の各列は、下側のテーブル534の列に対応する。上側のテーブル536は、Microsoft Accessに含まれる例示照会(query by example)(QBE)グリッドに類似する。述語は、上側のテーブル536の列に入力され、全ての述語はANDとして処理される1つの行に、およびORとして処理されるその間の行に入る。所定の照会を満足する結果は、フィルタボタン538を選択することによって、下側のテーブル534に示される。フィルタは、適切に標識されたボタン540、542、および544を使用することによって保存、削除、およびリセットされ得る。保存されたフィルタは、ドロップダウンリスト546を使用することによってロードされ得る。エクスポートボタン547を選択することにより、データをエクセルのスプレッドシートに書き込む。

【 0 0 3 6 】

さらなるユーザ照会を容易にするために、ユーザは、工程408において将来の照会のためのピボットフィールドとして使用されるべき新たなフィールドを特

定し得る。選択されたフィールドの要素は、新たなテーブルの列となる。図 5 F は、ピボット値が、ドロップダウンリスト 5 4 8 を使用することによっていかに選択されるかを示す。ピボット値は、下側のテーブル 5 3 4 の列に列挙される発現データを識別する。図 5 G は、ピボット列ドロップダウンリスト 5 5 0 が、下側のテーブル 5 3 4 の特定の列をピボットフィールドとして選択できるようにすることを示す。選択された列の入力は、左のリストボックス 5 5 2 に示され、そして右側のリストボックス 5 5 4 に移動して、それらをピボット化された表の行として含める。ユーザは、矢印キー 5 5 6 を選択して、右側のリストボックス 5 5 4 のアイテムを追加および削除する。ピボット操作を行うために、ユーザはピボットボタン 5 5 8 を選択する。

【 0 0 3 7 】

図 5 H は、ピボット操作の結果として示される組織タイプをフィルタするためのユーザインターフェースを示す。下側のテーブル 5 3 4 は、図 5 F ~ 5 G に参照して記載されるピボット操作の結果を示す。

【 0 0 3 8 】

上側のテーブル 5 3 6 は、ここで工程 4 1 0 において、異なる組織タイプから得られた実験の結果を使用して遺伝子をフィルタするための照会を特定するために使用される。再び、行における述語は AND として処理される。行の間の述語は、OR として処理される。照会を適切に定式化することによって、ユーザは、どの遺伝子が正常組織中でアップレギュレートされるか、および疾患組織中でダウンレギュレートされるかというような質問に答えることができる。示された *Entrez* 定義列には、パブリックドメインである *Entrez* データベースからの定義列を含む。「like "growth"」と記された示された照会は、ストリング「growth (増殖)」を、指定された列においてサブストリングとして有するレコードを保持する。

【 0 0 3 9 】

示された照会を満足する 1 つの条件は、遺伝子が、実験 4 0 0 2 7 3 6 D において 1 0 よりも大きい発現レベルを有すること、および実験 4 0 0 3 2 2 8 A において 1 0 よりも大きく、かつ実験 4 0 0 2 7 3 6 D における発現レベルの 0 .

6 倍よりも小さい発現レベルを有することである。照会を満足する代替の条件は、実験 4 0 0 2 7 3 6 D における発現レベルが 1 0 よりも大きく、そして実験 4 0 0 3 2 2 8 A における発現レベルが 1 0 よりも大きく、かつ実験 4 0 0 2 7 3 6 D における発現レベルの 1. 4 倍よりも大きいことである。

【 0 0 4 0 】

この照会は、実験 4 0 0 3 2 2 8 A と実験 4 0 0 2 7 3 6 D との間の特定倍の変化パターンを有する遺伝子を決定する。その実験間で有意な倍数の変化を有さない遺伝子がフィルタされて除かれる。詳細には、これは、実験 4 0 0 3 2 2 8 A の発現レベルが、実験 4 0 0 2 7 3 6 D の発現レベルの 6 0 % 未満であるか、または実験 4 0 0 3 2 2 8 A の発現レベルが、実験 4 0 0 2 7 3 6 D の発現レベルの 1 4 0 % よりも大きい全ての遺伝子を見い出す。両方の実験はまた、1 0 よりも大きい発現レベルを有するように制約される。

【 0 0 4 1 】

フィルタは、ボタン 5 6 0 および 5 6 2 をそれぞれ選択することにより保存またはリセットされ得る。下側のテーブル 5 3 4 に示されるレコードは、任意の列に分類され、そして列は、より見やすいように隠されるか、フリーズされるか、または再配置され得る。下側のテーブル 5 3 4 はまた、エクスポートボタン 5 6 4 をクリックすることによって、異なるフォーマットで保存され得、それには M i c r o s o f t E x c e l のようなスプレッドシートフォーマットが含まれる。保存されたフィルタは、プルダウンメニュー 5 6 6 を介してアクセスされ得るか、または削除ボタン 5 6 8 を選択することによって削除され得る。任意の遺伝子に関するさらなる情報は、その行をダブルクリックすることによって得ることができる。これは、インターネットブラウザプログラムをロードし、そして遺伝子に関する情報を記憶する E n t r e z ウェブサイトのようなウェブサイトを開く。次いで、そのブラウザプログラムは、その遺伝子の入力を表示する。

【 0 0 4 2 】

工程 4 1 2 で、グラフボタン 5 7 0 を選択することによって、ユーザは、図 5 1 に示すスキャッタープロット表示 5 7 2 を呼び出す。2 つの実験が、ドロップダウンリスト 5 7 4 および 5 7 6 をそれぞれ x 軸および y 軸について使用する比

較のために選択されている。グラフは、スキャッター構築ボタン578を選択することによって選択することによって作成される。スキャッタープロット上の各点は、特定の遺伝子に対応する。点は、両方の実験におけるその測定された発現レベルにしたがってグラフ上に配置される。ボックス580をチェックすることによって、ユーザは、その遺伝子が、その実験の両方(2P)、一方(1P)に存在するか、またはいずれにも存在しない(OP)か否かにより、その点が色でコード付けられるように選択し得る。ボックス582の1つ以上をチェックすることによって、ユーザは、この分類による遺伝子を示すかまたは示さないかを選択することができる。

【0043】

ボックス584において適切な選択をすることによって、ユーザは、将来のマウスクリックについての解釈を選択し得る。このシステムについての1つの選択は、マウスクリックに対して何もしないことである。このシステムに対する別の選択は、マウスクリックによって選択した点についての遺伝子データを示すことである。その遺伝子データは、ボックス586に現れ、それにはアクセス番号、遺伝子名、種々の実験において測定された発現レベル、および各実験に対する発現呼び出し(非存在または存在のいずれか)が含まれる。Entrez定義名もまた示される。入力をダブルクリックすることによって、インターネットブラウザが立ち上がり、その遺伝子についてのEntrez入力を示す。

【0044】

ユーザはまた、ボックス584中の「rope(縄で囲う)」を選択して、比較のために興味深い点を多角形でそれらを取り囲むことにより収集し得る。棒グラフに含めるべき遺伝子を囲む各マウスクリック間で、直線が自動的に引かれる。ユーザは、棒グラフをボタン588を選択することによって表示し得る。

【0045】

工程414で、図5Jは、図5Iのスキャッタープロットにおいて囲まれた遺伝子についての棒グラフ590を示す。図5Jにおける棒の各グループ化は、遺伝子に対応する。グループ化内での各棒は、1つの実験に対応し、そして説明592にしたがって色でコードづけられる。最初に2つのみの実験が表示され、そ

の2つの実験は、図5 Iのスカッタープロットの軸に対応する。しかし、ユーザは、ボックス594から実験をさらに選択し得る。一旦、所望の実験が選択されると、ユーザは、構築ボタン596を選択し、所望の棒グラフを表示する。テーブル598は、示された遺伝子の発現レベルを示す。

【 0 0 4 6 】

図5 Jのディスプレイについて、ボックス600においてオプション「gene (遺伝子)」が選択される。各遺伝子についての発現レベルの個々のプロット(それらは実験にわたって変化するため)を見るために、ユーザは、構築ボタン596を選択する前に、ボックス600においてオプション「experiment (実験)」を選択し得る。これにより、図5 Kに示されるように、折れ線グラフ602が生成される。その実験は、水平軸に沿ってボックス594に特定される順序でアレンジされる。各遺伝子は、それが実験にわたって変化するので、その発現レベルに対応するそれ自体のトレースを有する。説明604は、各遺伝子についてのトレースを識別する。実験の位置を水平軸に沿って変化させるために、ユーザは、上向きおよび下向き矢印606および608を使用して、その位置を変化させる。この特徴により、さらなる配列決定知識を反映させるために実験を記録することが可能となる。例えば、その実験が疾患または処置の進行のような時間経過を表す場合、それらは時間系列でグラフにより整列され得る。次いで、そのグラフは、選択された遺伝子についての時間の関数として発現レベルにおける変化を表す。スライダーアイコン612は、折れ線グラフ602がスクリーンに収まりきらない場合、ユーザが水平軸に沿ってスクロールすることを可能にする。マーカーチェックボックス614は、特定の発現レベルを規定する折れ線グラフ602を横切る水平線を示す。これにより、ユーザは選択されたレベルより上のデータ点を容易に見ることができる。

【 0 0 4 7 】

遺伝子に関するより多くの情報が、グループ内の任意のバーをクリックすることによって得ることができる。その遺伝子の情報の全てが、図5 Lに示されるように、別個のウィンドウ610に表示される。

【 0 0 4 8 】

上記の明細書において、本発明はその特定の例示的な実施態様を参照して記載されてきた。しかし、種々の修飾および改変が、添付の請求項に示される本発明のより広い精神および範囲ならびにその等価物の全範囲から逸脱することなく、それに対して成され得ることが明らかである。例えば、「発現レベル」が言及されるところではどこでも、任意の化合物のその測定された濃度を当業者は置換し得ることが理解される。また、「遺伝子」が言及されるところではどこでも、当業者は、その用語を「発現配列タグ」と置換し得る。

【図面の簡単な説明】

【図 1】

図 1 は、本発明のソフトウェア実施態様を実行するために使用され得るコンピュータシステムの一例を示す。

【図 2】

図 2 は、代表的なコンピュータシステムのシステムブロック図を示す。

【図 3】

図 3 は、本発明の 1 つの実施態様に従って発現データを展開する工程を記載するフローチャートである。

【図 4】

図 4 は、本発明の 1 つの実施態様に従って発現データベースを照会する工程を記載するフローチャートである。

【図 5】

図 5 A ~ 5 L は、本発明の 1 つの実施態様に従って発現データベースを照会するためのユーザインターフェースを示す。

【 図 1 】

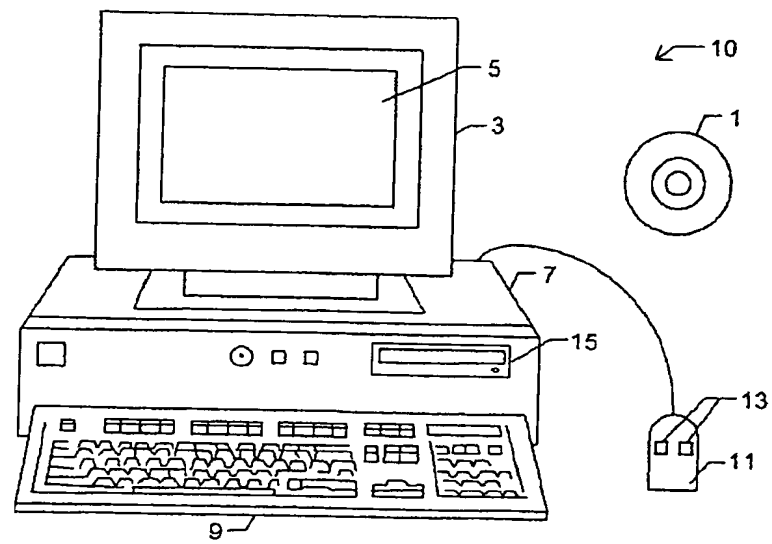
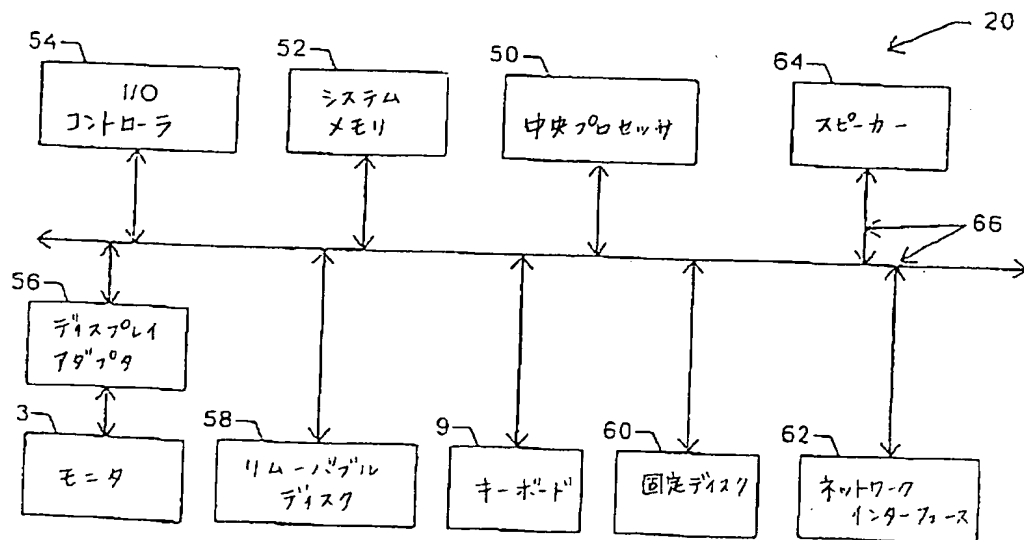
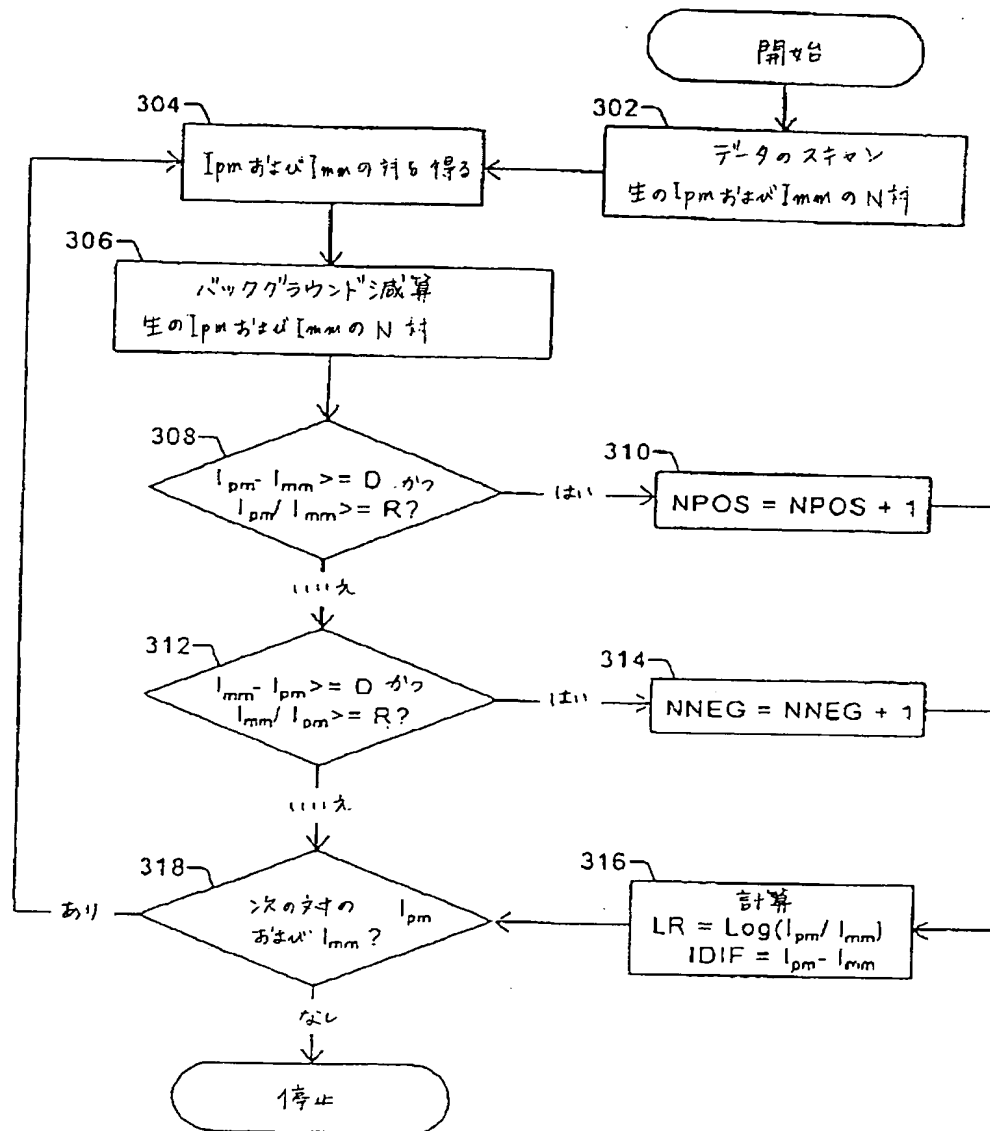


FIG. 1

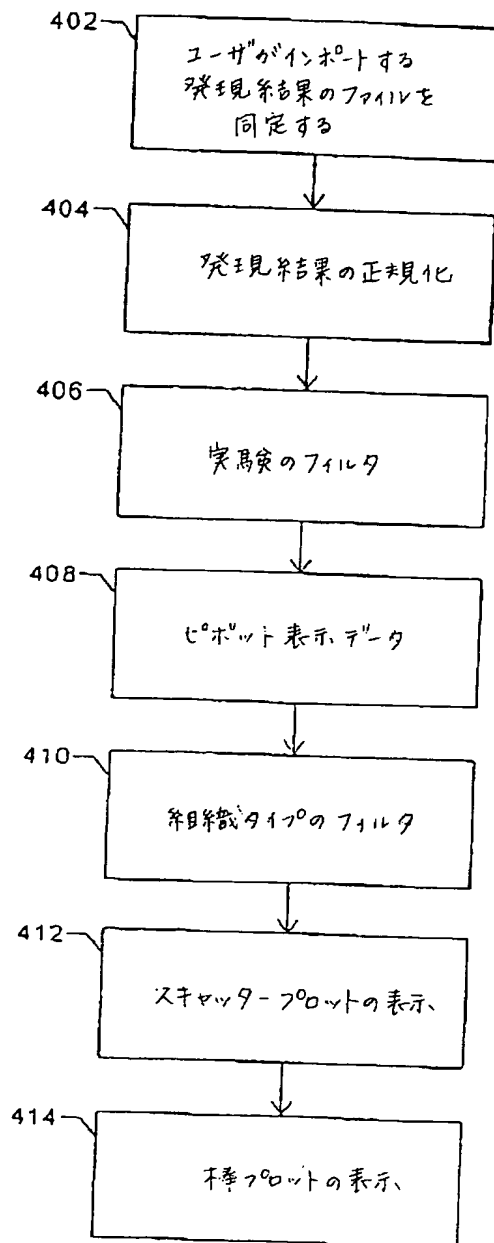
【 図 2 】



【 図 3 】



【 図 4 】



【 図 5 A 】

Import and normalization

504

506

512

508

514

502

Gene Chip Results

FileName	Code	Replicate	ChipDesignCode	Sam
g:\genexel\Gish Data\text files from CHP files\sy050103.txt	4002736D	0 A		cand
g:\genexel\Gish Data\text files from CHP files\sy050201.txt	4002736D	0 B		cand
g:\genexel\Gish Data\text files from CHP files\sy050601.txt	4002736D	0 C		cand
g:\genexel\Gish Data\text files from CHP files\sy050701.txt	4002736D	0 D		cand
g:\genexel\Gish Data\text files from CHP files\sy080501.txt	4003228A	0 A		cand
g:\genexel\Gish Data\text files from CHP files\sy080601.txt	4003228A	0 B		cand
g:\genexel\Gish Data\text files from CHP files\sy080701.txt	4003228A	0 C		cand
g:\genexel\Gish Data\text files from CHP files\sy080801.txt	4003228A	0 D		cand
g:\genexel\Gish Data\text files from CHP files\sy052001.txt	4003265J	0 A		cand

select files

Analysis Type

Species

File List

Human

Kurt's Famous 48

delete file list

save file list

Import

reset file list

Gene Expression Levels

FIG. 5A

FIG. 5B

【 図 5 C 】

Import and normalization

519

518

Import Normalization

select files

Analysis Type

Species

File List

Human

Animal

Ecoli

Human

Mouse

Yeast

Gene Chip Results

FileName	ChipDesignCode	Sample
G:\Genexel\Gish Data\text files from CHP files\sy050201.txt	4002736D	C
G:\Genexel\Gish Data\text files from CHP files\sy050701.txt	4002736D	A
G:\Genexel\Gish Data\text files from CHP files\sy080701.txt	4003228A	B
G:\Genexel\Gish Data\text files from CHP files\sy080501.txt	4003228A	D
G:\Genexel\Gish Data\text files from CHP files\sy080601.txt	4003228A	C
G:\Genexel\Gish Data\text files from CHP files\sy080801.txt	4003228A	A
G:\Genexel\Gish Data\text files from CHP files\sy050803.txt	4003265J	B
G:\Genexel\Gish Data\text files from CHP files\sy050803.txt	4003265J	D
G:\Genexel\Gish Data\text files from CHP files\sy050803.txt	4003265J	C

delete file list

save file list

reset file list

Import

Gene Expression Levels

FIG. 5C

【 図 5 D 】

Import and normalization

Import
Normalization

Chip Variability

Gene Name	Accession #
HUMGAPDH/M33197	HUMGAPDH/M33197
HSAC07/X00351	HSAC07/X00351
*	

Scale Value

☐ Use
 ☒ Don't Use

apply

Tissue Variability

Scale Value

☐ Use
 ☒ Don't Use

Transform

Negative to log

☒ Use
 ☐ Don't Use

reset

Gene Expression Levels

FIG. 5D

【 図 5 E 】

532 filter experiments 545 Abs call 544 delete filter

536 MM Excess Pos/Neg Avg Diff Abs Call Code TissueType Chip Access

538 filter reset filter save filter export close

542

534

Experiment Na	Gene Name	Positive	Negative	Pairs	Used	Pairs	inAvg	Pos FA
sy050103	100N95678	8.00	3.00	20.00	20.00	18.00	18.00	
sy080501	100N95678	12.00	1.00	20.00	20.00	18.00	18.00	
sy062001	100N95678	8.00	2.00	20.00	20.00	18.00	18.00	
sy071501	100N95678	12.00	2.00	20.00	20.00	18.00	18.00	
sy082001	100N95678	10.00	0.00	20.00	20.00	17.00	17.00	
sy080503	100N95678	11.00	1.00	20.00	20.00	18.00	18.00	
Record		89748	41					

Pivot Value: FixedAvgDiff Pivot Column: pivot

FIG. 5E

[illegible]

FIG. 5F

【 図 5 G 】

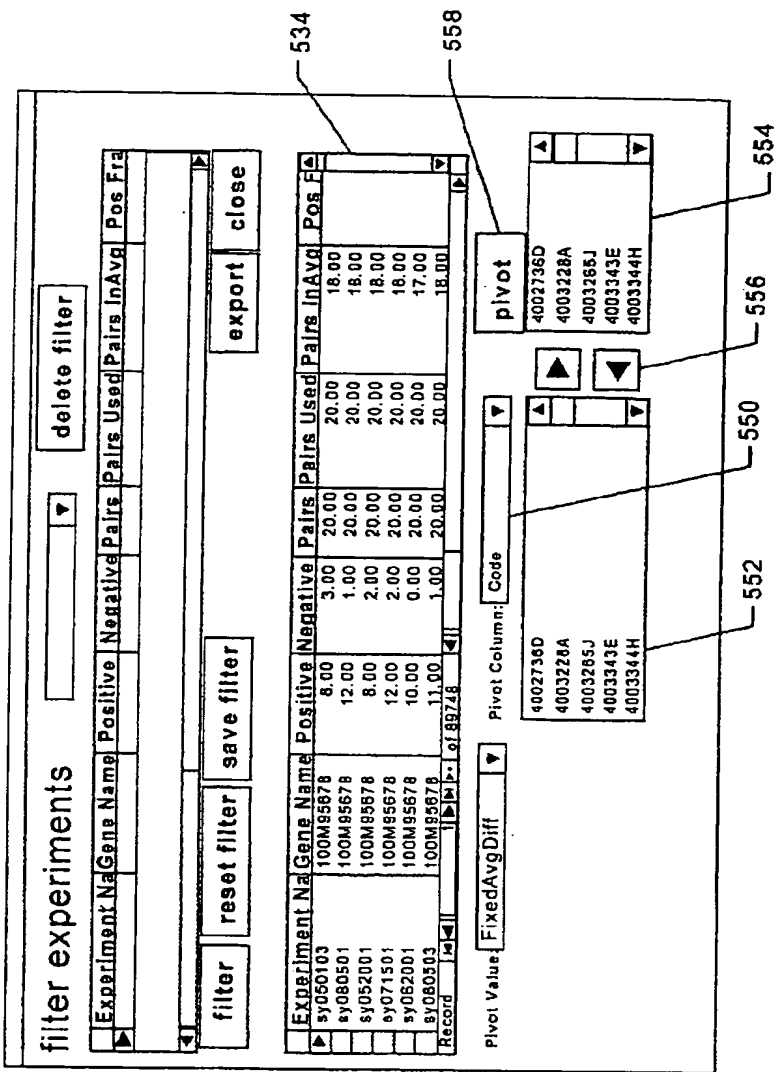


FIG. 5G

【 図 5 H 】

filter tissue types

Gene Name 4002736D 4003228A 4003265J 4003343E 4003344H 4003353C 4003353D

Entrez Definition like "growth" like "growth" like "growth" like "growth" like "growth" like "growth"

>10 and <0.6' 4002736D like "growth"

>10 and >1.4' 4002736D like "growth"

renew with growth

delete filter

filter reset filter save filter 570 graph export close

536 562 564 534

Gene Name	4002736D	4003228A	4003265J	4003343E	4003344H	4003353C	4003353D
1000M86578	424.14	371.34	220.06	301.90	258.31	378.82	378.82
1000T51617	10.59	15.85	0.06	19.48	33.87	31.72	31.72
1001M83363	10.36	55.88	0.03	34.66	34.81	57.97	57.97
1002R06239	974.83	141.27	62.31	80.49	103.63	88.36	88.36
1003L14922	1.23	27.92	11.11	1.69	0.08	0.09	0.09
1003M32402	0.04	0.06	0.06	0.08	0.07	0.05	0.05
1004T51844	61.38	18.65	15.25	15.38	22.07	5.83	5.83
1004L02932	0.03	13.88	0.30	8.74	0.04	1.18	1.18
1004T51849_J	350.21	184.89	134.93	237.75	137.98	334.93	334.93
1004T51849_I	297.03	0.02	0.01	84.67	11.77	0.00	0.00
10046T51849_I	224.27	0.01	0.01	81.43	0.01	0.00	0.00
1005L07592	71.98	119.08	86.34	118.06	97.49	167.91	167.91
Record	1 of 7454						

FIG. 5H

【 図 5 I 】

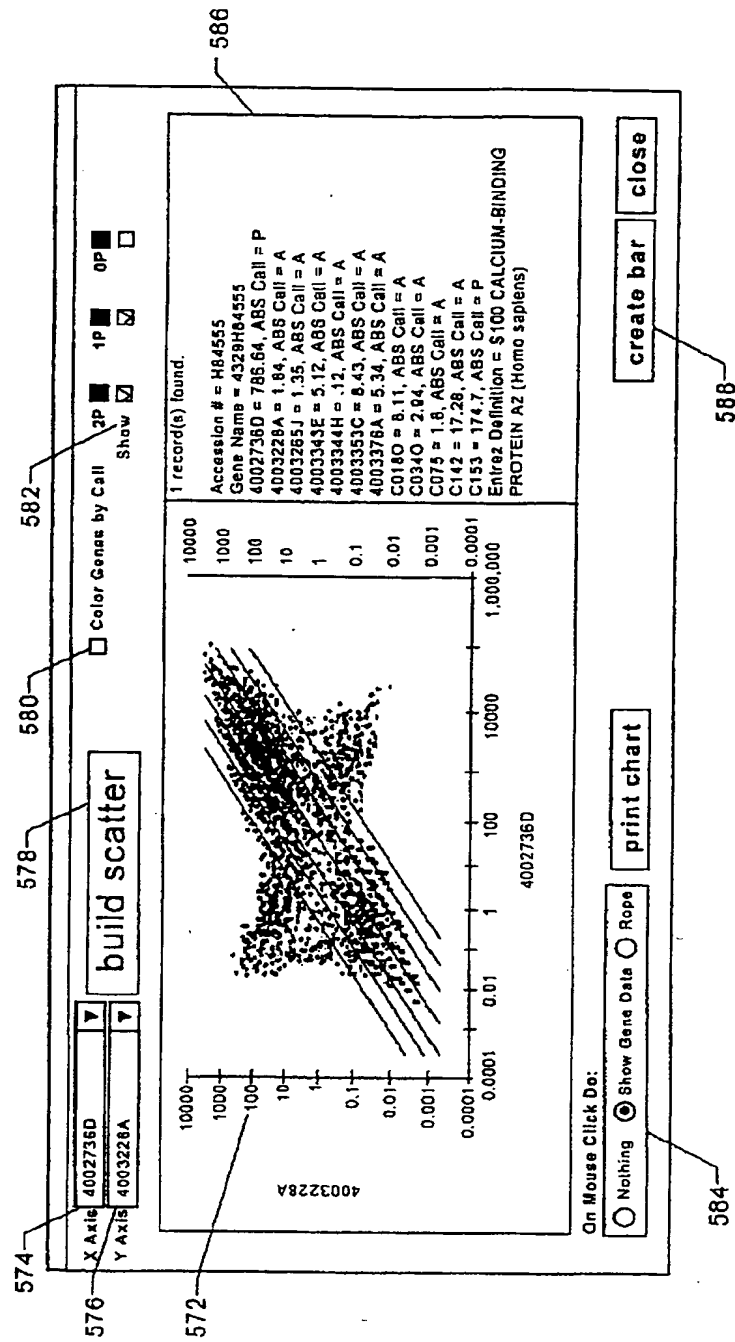


FIG. 5I

【 図 5 J 】

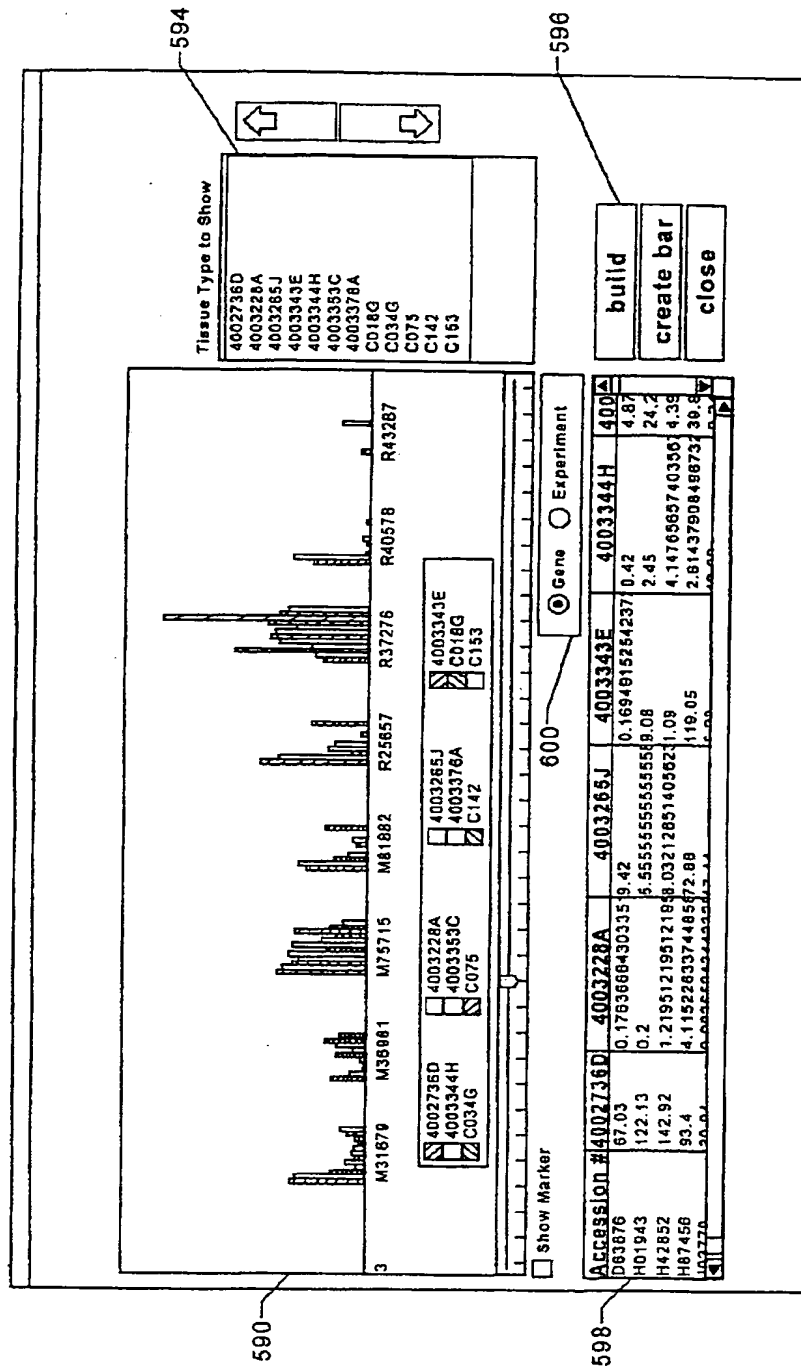


FIG. 5J

【 図 5 K 】

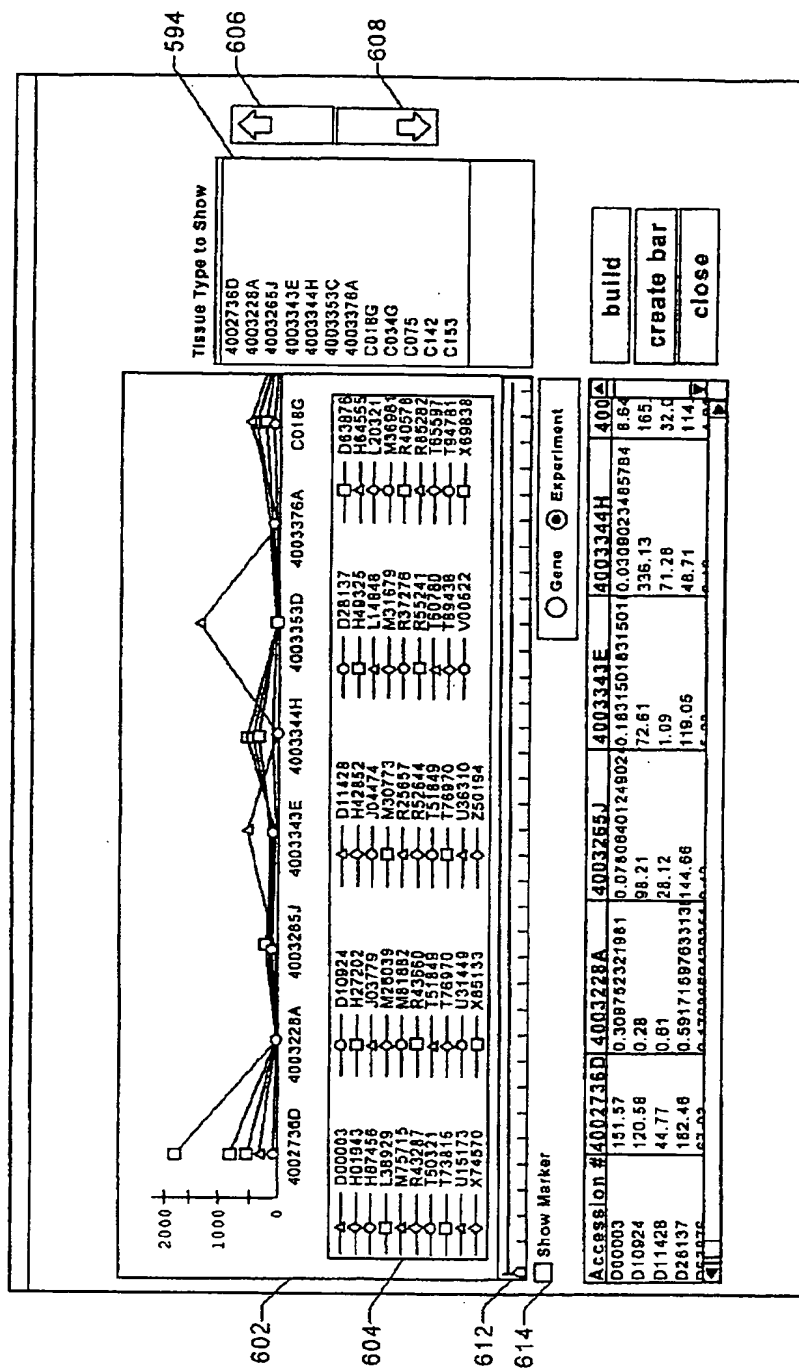


FIG. 5K

【 図 5 L 】

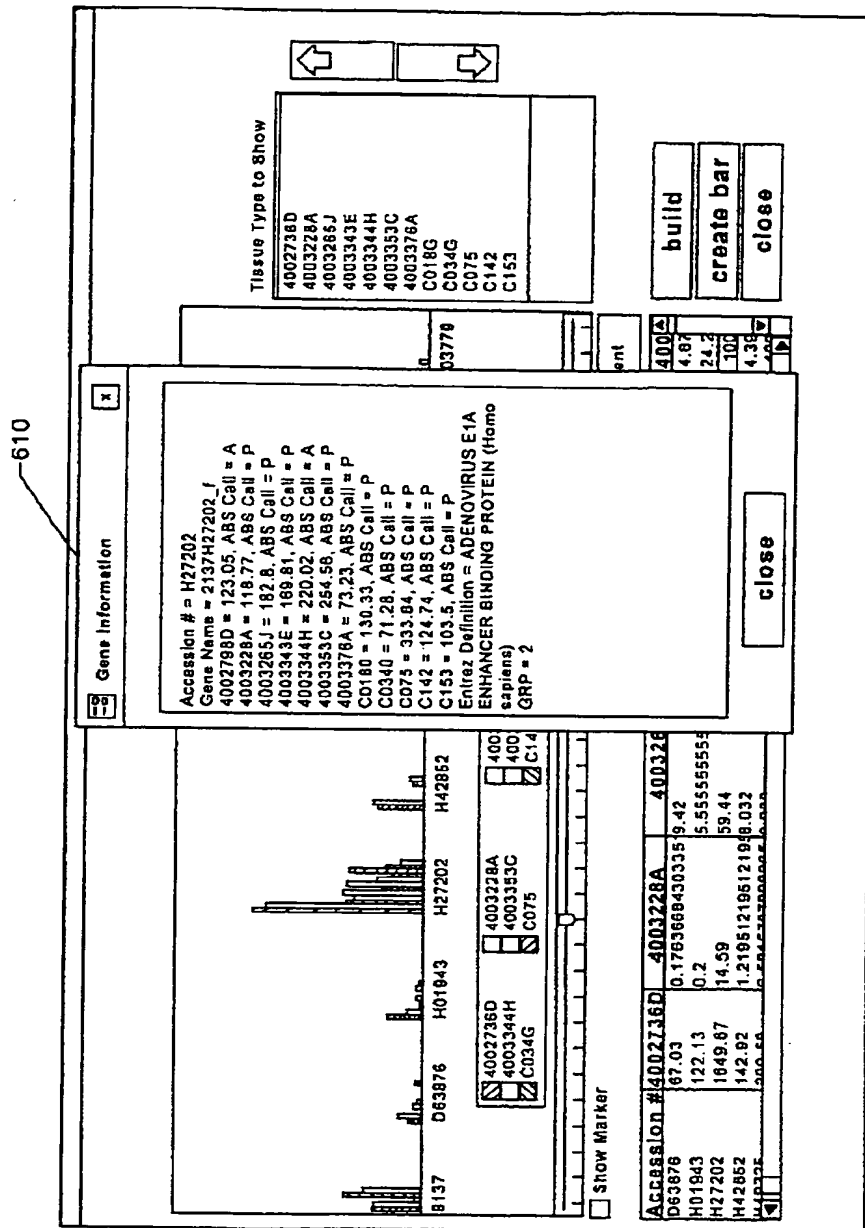



FIG. 5L

[国際調査報告]

INTERNATIONAL SEARCH REPORT		International application No. PCT/US98/15151
A. CLASSIFICATION OF SUBJECT MATTER IPC(6) : C12Q 1/68; G06F 17/30, 159/00 US CL : 435/6; 707/1; 128/923 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/6; 707/1; 128/923 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS, STN. Search terms: transcript, transcripts, cdna#, mma#, est#, frequenc?, abundanc?, distribut?, computer#, algorithm#		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	OKUBO et al. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression, Nature Genetics. November 1993, Vol. 2, No. 3, pages 173-179, see entire document.	1, 2, 4, and 8-10 3, 5-7, and 11-27
Y	IntelliGenetics Suite (TM), Release 5.4, Advanced Training Manual, January 1993, published by IntelliGenetics, Inc., 700 East El Camino Real, Mountain View, California 94040, USA, pages (1-6) - (1-19) and (2-9) - (2-14), see entire document.	3, 5-7, and 11-27
X - Y	WO 96/23078 A1 (INCYTE PHARMACEUTICALS, INC.) 01 August 1996, see entire document.	1-10 and 13-27 11 and 12
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *A* document defining the general state of the art which is not considered to be of particular relevance *X* document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *B* earlier document published on or after the international filing date *Y* document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combinations being obvious to a person skilled in the art *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *A* document member of the same patent family	
Date of the actual completion of the international search 16 OCTOBER 1998	Date of mailing of the international search report 29 OCT 1998	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer  JAMES MARTINELL Telephone No. (703) 308-0196	

フロントページの続き

- (31)優先権主張番号 60/069, 436
(32)優先日 平成9年12月11日(1997. 12. 11)
(33)優先権主張国 米国(US)
(81)指定国 EP(AT, BE, CH, CY,
DE, DK, ES, FI, FR, GB, GR, IE, I
T, LU, MC, NL, PT, SE), JP